

NEXUS

OpenCitations

an infrastructure for
open bibliographical
metadata

Ivan Heibi

University of Bologna

Outline

- OpenCitations
 - Overview
 - Citations
 - OpenCitations Indexes
 - Integration with Nexus services
- Recent releases
 - DOCI and POCI
 - META
 - Indexed resources in OC
- Querying and usage
 - SPARQL endpoints
 - REST APIs
 - OpenCitations access token
 - Dumps

OpenCitations

<https://opencitations.net/>

Overview

OpenCitations (<http://opencitations.net>) is an independent infrastructure organization

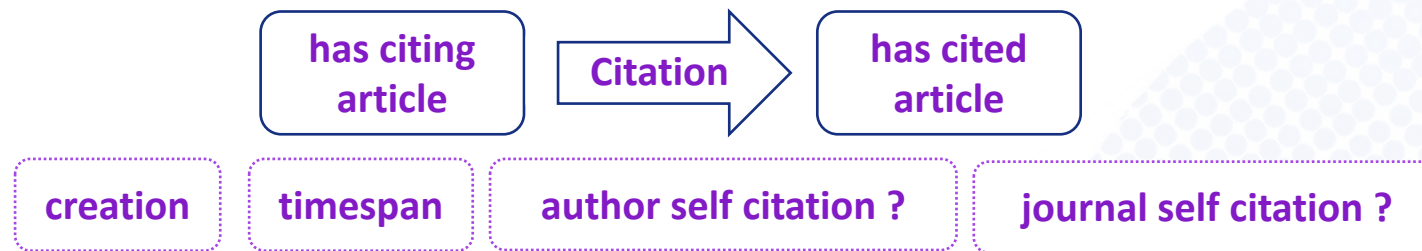
- dedicated to open scholarship and the publication of open bibliographic and citation data by the use of Semantic Web technologies
- engaged in advocacy for open citations and open bibliographic metadata, particularly via I4OC (<https://i4oc.org>) and I4OA (<https://i4oa.org>)

It provides:

- a data model: the OpenCitations Data Model (based on the SPAR Ontologies)
- bibliographic and citation data (CC0)
- software: in our GitHub repository, released with open source licenses
- online services: REST APIs, SPARQL endpoints, dumps and interfaces

Citations in OpenCitations

Citation data should be structured, separable, and open. In OpenCitations citations are hosted as Linked Open Data, defined using the [OpenCitations Data Model \(OCDM\)](#) as first-data entities



Advantages:

- all the information regarding each citation is available in one place.
- citations become easier to describe, distinguish, count and process.

Citations are identified with the [Open Citation Identifier \(OCI\)](#), a globally unique persistent identifier (PID) for the identification of open bibliographic citations stored in a specific database or in other kinds of storages.

OpenCitations Indexes

COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations –

an RDF dataset containing data of all the citations that are specified by the open references to DOI-identified works present in Crossref

CROCI, the Crowdsourced Open Citations Index – individuals (identified by ORCID) may deposit citation information that they have a legal right to submit, and within which these submitted citation data will be published under a CC0 public domain waiver

At the end of November 2022 OpenCitations had:

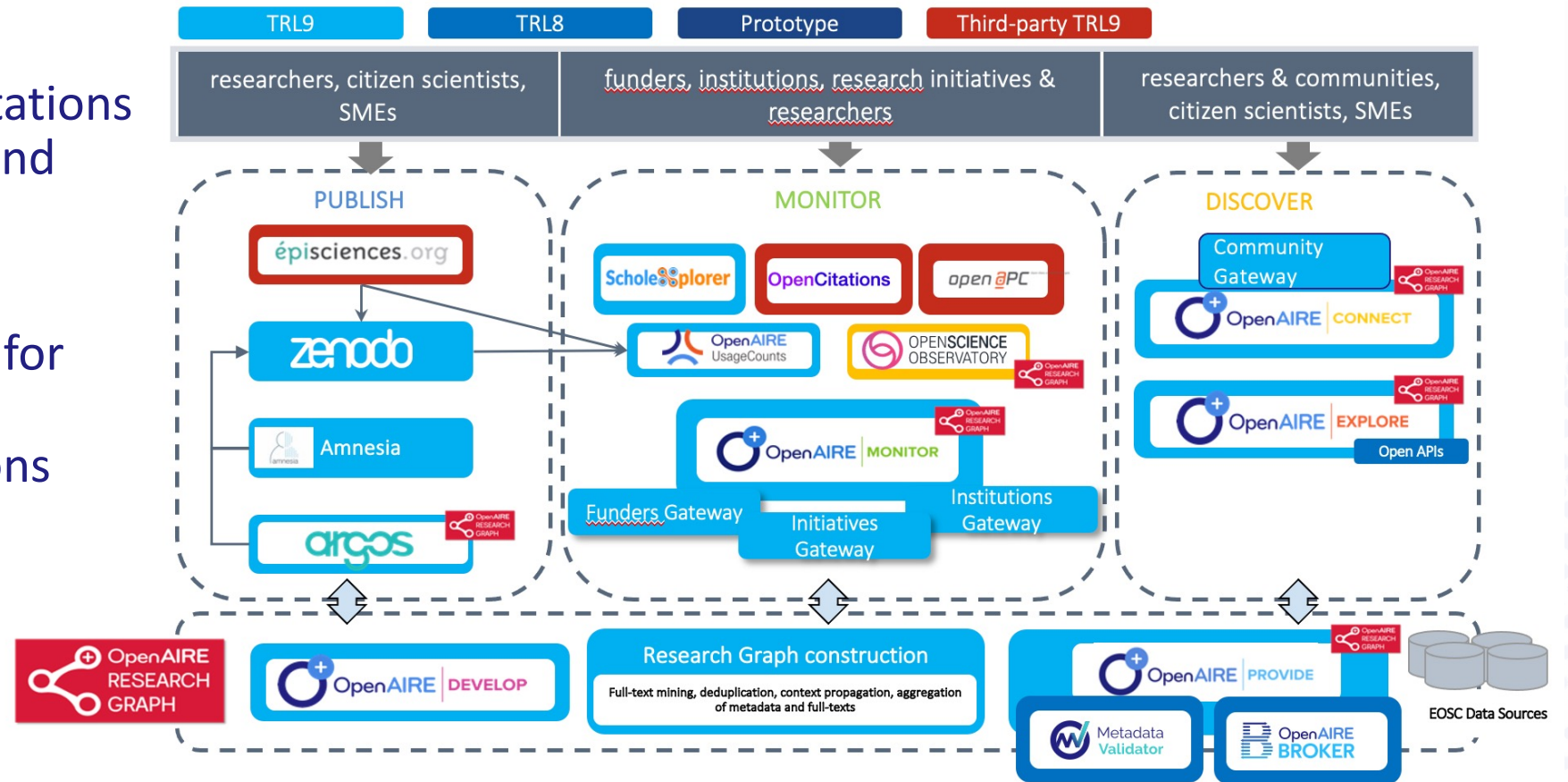
1,363,718,366 citation links

75,030,924 bibliographic resources

Integration with Nexus services

Research Graph

- ingestion of OpenCitations data (bibliographic and citation data) to the Research Graph
- software developed for enabling ingestion to/from OpenCitations



Recent releases

Data and Services released on December 2022

DOCI

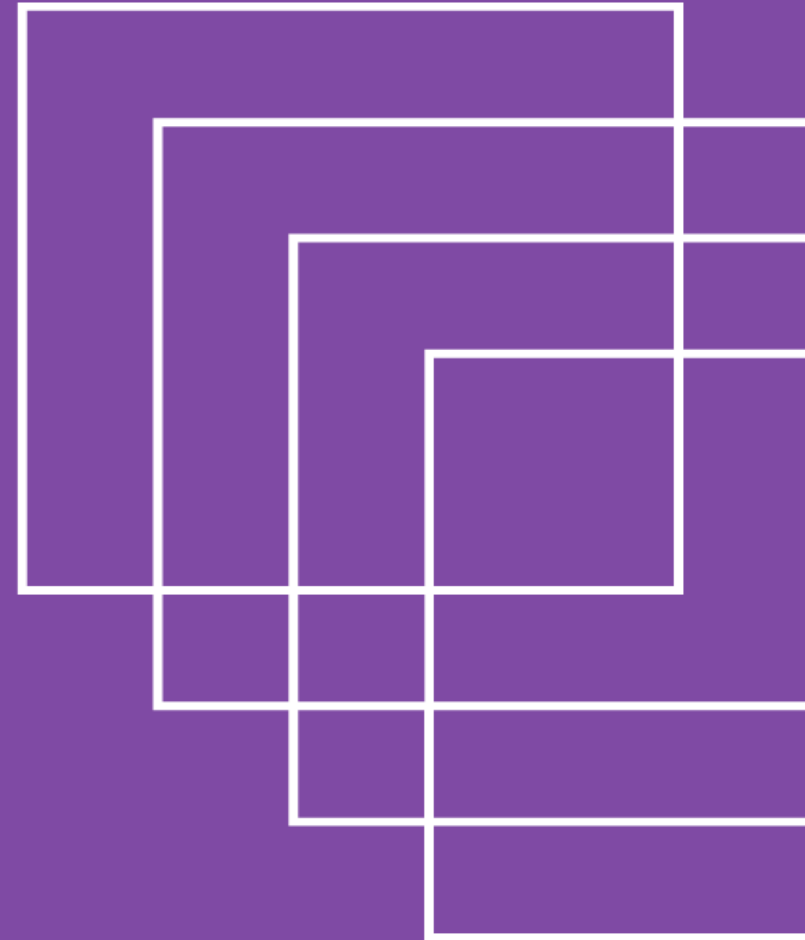
OPEN CITATIONS FROM DATACITE

169,822,752

CITATIONS

1,753,858

**BIBLIOGRAPHIC
RESOURCES**



POCI

OPEN CITATIONS FROM PUBMED

717,654,703

CITATIONS

29,005,551

**BIBLIOGRAPHIC
RESOURCES**

OPENCITATIONS META

87,321,593

**BIBLIOGRAPHIC
RESOURCES**

277,750,235

AUTHORS

710,226

VENUES

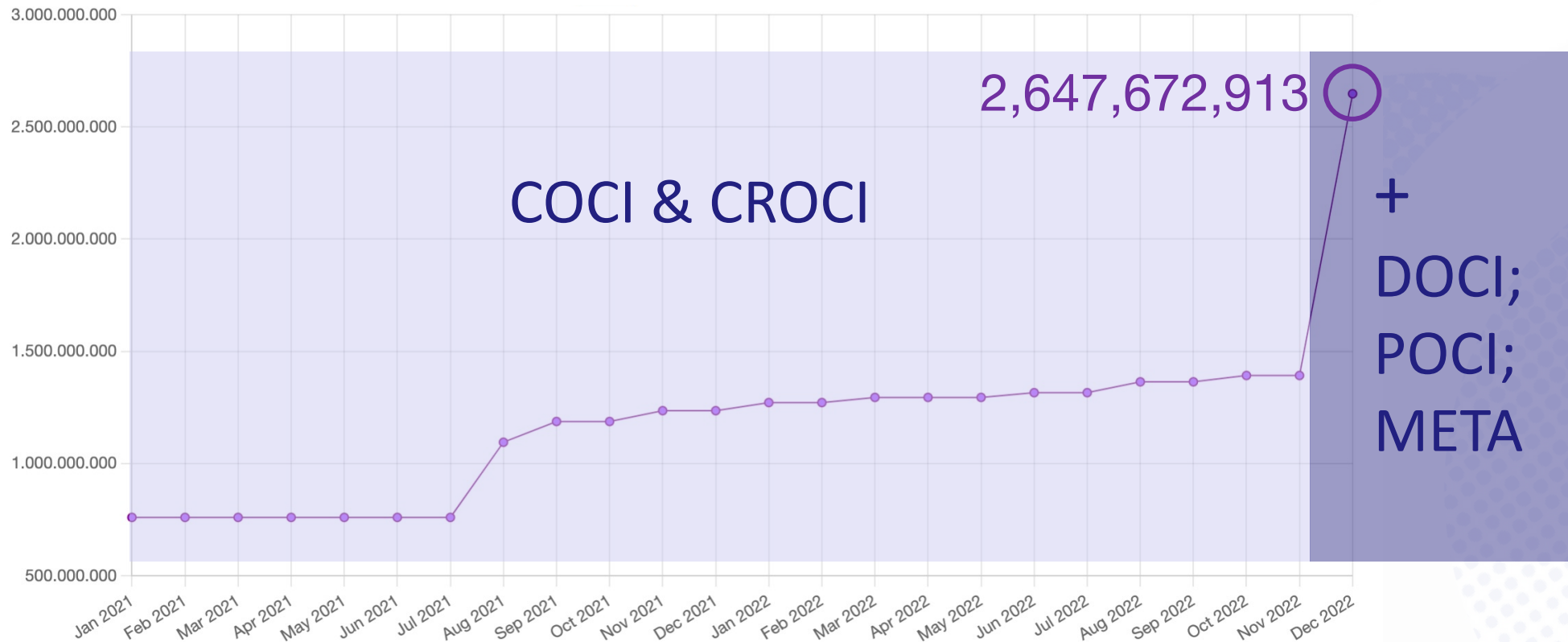
2,359,301

EDITORS

17,268

PUBLISHERS

Indexed resources in OpenCitations



Querying Data

<https://opencitations.net/querying>

SPARQL endpoints

OpenCitations maintains a SPARQL endpoint for all its datasets. When such a SPARQL endpoint is accessed with a browser, it shows an editor GUI generated with YASGUI

Requires technical background on semantic web technologies and SPARQL language

OpenCitations Indexes SPARQL endpoint

```
1 PREFIX cito:<http://purl.org/spar/cito/>
2
3 SELECT DISTINCT ?citing_entity ?cited_entity ?creation_date ?timespan WHERE {
4   GRAPH <https://w3id.org/oc/index/coci/> {
5     ?citation a cito:Citation ;
6             cito:hasCitingEntity ?citing_entity ;
7             cito:hasCitationCreationDate ?creation_date ;
8             cito:hasCitationTimeSpan ?timespan ;
9             cito:hasCitedEntity ?cited_entity
10    }
11 } LIMIT 10
```



Raw Response

Table

Pivot Table

Google Chart

Search:

Show 50

entries

REST APIs

Data in any of the OpenCitations datasets can be retrieved by using an HTTP REST API. This solution represents a convenient access to the data included in the OpenCitations datasets for Web developers and users who are not necessarily experts in Semantic Web technologies.

Implemented by means of RAMOSE, the Restful API Manager Over SPARQL Endpoints

(<https://github.com/opencitations/ramose>)

The unifying REST API for all the OpenCitations Indexes

DESCRIPTION

PARAMETERS

OPERATIONS

/references/{doi}

/citations/{doi}

/citation/{oci}

/metadata/{dois}

/citation-count/{doi}

/reference-count/{doi}

HOME

THE UNIFYING REST API FOR ALL THE OPENCITATIONS INDEXES

VERSION: Version 1.1.0 (2020-03-25)

API URL: <https://w3id.org/oc/index/api/v1>

CONTACT: contact@opencitations.net

LICENSE: This document is licensed with a [Creative Commons Attribution 4.0 International License](#), while the REST API itself has been created using [RAMOSE](#), the *Restful API Manager Over SPARQL Endpoints* created by [Silvio Peroni](#), which is licensed with an [ISC license](#).

DESCRIPTION

[↑](#) BACK TO TOP

This document describe the REST API for accessing the data stored in all the [OpenCitations Indexes](#) hosted by [OpenCitations](#). This API implements operations to retrieve the citation data for all the references to other works appearing in a particular bibliographic entity, or the citation data for all the references appearing in other works to a particular bibliographic entity, given the DOI of a bibliographic entity, or to retrieve citation data about a particular citation identified by means of its [Open Citation Identifier \(OCI\)](#).

OpenCitations access token

Before accessing our services, we encourage getting the OpenCitations Access Token.

An opaque string that anonymously identifies a unique user of the OpenCitations' APIs.

Its usage is not compulsory; however, it will help OpenCitations incredibly, by enabling us to monitor the number of the unique users accessing our data and services.

<https://opencitations.net/accesstoken>

OpenCitations Access Token

Email:

E.g. user@example.com

Get token



Non sono un robot



reCAPTCHA
Privacy - Termini

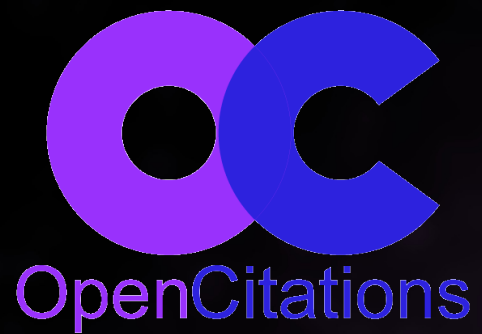
Dumps

Data dumps of the OpenCitations Indexes – COCI, CROCI, DOCI, PO CI, and META. Dumps are available on Figshare (<https://figshare.com/>) in CSV, N-Triples and Scholix formats.

This option is highly recommended for massive and large data analysis on the resources of OpenCitations

<https://opencitations.net/download>

Type and format	Archive	Size
Citation data (CSV)	ZIP	238.5 GB (37.5 GB zipped)
Citation data (N-Triple)	ZIP	1.6 TB (73.1 GB zipped)
Citation data (Scholix)	ZIP	1.3 TB (38.8 GB zipped)
Provenance data (CSV)	ZIP	330 GB (20 GB zipped)
Provenance data (N-Triple)	ZIP	3.3 TB (78 GB zipped)



THANKS

Email

Ivan.heibi2@unibo.it

Twitter

@ivanhb
